Features selection in Discrete Discriminant Analysis

Anabela Marques[1], Ana Sousa Ferreira[2] and Margarida Cardoso[3]

[1]ESTBarreiro, Setúbal Polytechnic, Portugal
(E-mail:anabela.marques@estbarreiro.ips.pt)
[2]LEAD, Faculty of Psychology, University of Lisbon, Portugal, CEAUL and UNIDE
(E-mail:asferreira@fp.ul.pt)
[3]Dep. of Quantitative Methods of ISCTE – Lisbon University Institute, Portugal, UNIDE
(E-mail:margarida.cardoso@iscte.pt)

Abstract. In discrete discriminant analysis dimensionality problems occur, particularly when dealing with data from the social sciences, humanities and health. In these domains, one often has to classify entities with a high number of explanatory variables when compared to the number of observations available.
In the present work we address the problem of features selection in classification, aiming to identify the variables that most discriminate between the *a priori* defined classes, reducing the number of parameters to estimate, turning the results easier to interpret and reducing the runtime of the methods used. We specially address classification using a recently methodological approach based on a linear combination of the First-order Independence Model (FOIM) and the Dependence Trees Model (DTM).
Data of small and moderate size are considered.
Keywords: Discrete Discriminant Analysis; Combining models; Dependence Trees model; First Order Independence model; Hierarchical Coupling procedure; Variable Selection.

References

1. Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, **1**, 289-300, 1995.
2. A. P. Duarte Silva. Classificação supervisionada para dados de elevada dimensão. *Livro de resumos das XVII Jornadas de Classificação e Análise de Dados* (JOCLAD 2010), ISCTE Eds., 17, 2010.
3. A. Marques, A. Sousa Ferreira and M. Cardoso. Uma proposta de combinação de modelos em Análise Discriminante Discreta. *Estatística – Arte de Explicar o Acaso*, em Oliveira, I. *et al*. (Eds.), Ciência Estatística, Edições SPE, 393-403, 2008.
4. A. Sousa Ferreira. *A Comparative Study on Discrete Discriminant Analysis through a Hierarchical Coupling Approach*. In Classification as a Tool for Research, Studies in Classification, Data Analysis, and Knowledge Organization; Hermann Locarek-Junge, Claus Weihs (Eds.), Springer-Verlag, Heidelberg-Berlin, 137-145, 2010.